

Bayesian nonparametric estimation of multivariate survival functions

Alan Riva Palacio and Fabrizio Leisen

School of Mathematics, Statistics and Actuarial Sciences, University of Kent
Sibson Building, Canterbury, Kent CT2 7FS

Abstract

In many real problems, dependence structures more general than exchangeability are required. For instance, in some settings partial exchangeability is a more reasonable assumption. For this reason, vectors of dependent Bayesian nonparametric priors have recently gained popularity. They provide flexible models which are tractable from a computational and theoretical point of view. In this paper, we focus on their use for estimating multivariate survival functions. Our model extends the work of Epifani and Lijoi (2010) to an arbitrary dimension and allows to model the dependence among survival times of different groups of observations. Theoretical results about the posterior behaviour of the underlying dependent vector of completely random measures are provided. The performance of the model is tested on a simulated dataset arising from a distributional Clayton copula.

1 Introduction

In many real applications, exchangeability can be a too restrictive assumption. Consider, for instance, data from different related studies, such as in randomized clinical trials where subjects are enrolled at different study centers or hospitals. In this case, it is more reasonable to consider that the outcome of the treatment is exchangeable among patients within the same hospital instead of patients from different centers. Factors specific to each hospital might have significant influence on the outcome distribution. Still, we observe a dependence component among hospitals. Another example is when we want to model the word occurrences in newspaper articles, for classification purposes. It is reasonable to think that articles in the same topic are exchangeable (in terms of word frequency) but not across topics. Indeed, different topics tend to use diverse vocabularies. Still, there is a dependence component across topics, since some articles can be crossclassified. For instance, an article about French wine could also refer to history, travels and weather.

The Bayesian nonparametric literature recently focused on proposing new models for dealing with data which exhibits a sort of local exchangeability within each group and, at the same time, a dependence among the groups, see [12] and the references therein. For instance, in survival analysis, [21] introduced a new class of vectors of random hazard rate functions that are expressed as kernel mixtures of dependent completely random measures.

Survival analysis is a branch of statistics which focuses on the modelling of time duration until one or more events occur such as death of a biological organism or failure in mechanical systems, see [1]. The main object of interest in survival analysis is the *Survival Function* $S(t)$ which measures the probability of surviving after time t . The analysis of survival data has been one of the first areas of application of Bayesian nonparametric techniques, see for instance [6], [11], [7]. The classical approach of [6] consists on defining the survival function of a lifetime T as follows

$$S(t) = \mathbb{P}[T > t | Y] = \mathbf{e}^{-Y(t)}, \quad (1)$$

where $Y(t)$ is a real valued stochastic process which has independent increments. Furthermore, it is assumed that $Y(t)$ is almost surely (a.s) non-decreasing, right continuous and such that

$$\lim_{t \rightarrow -\infty} Y(t) = 0 \quad \text{and} \quad \lim_{t \rightarrow \infty} Y(t) = \infty.$$

Under these assumptions the cumulative distribution functions $F(t) = 1 - S(t)$ is *neutral to the right* (NTR), namely a random cumulative distribution function such that the normalized increments,

$$F(t_1), (F(t_2) - F(t_1))/(1 - F(t_1)), \dots, (F(t_k) - F(t_{k-1}))/ (1 - F(t_{k-1})),$$

are independent for every $t_1 < \dots < t_k$. In the present work, we focus on processes $Y(t)$ built on *completely random measures* (CRM's), which are random measures with the property of inducing mutually independent random variables when evaluated in disjoint sets (see Section 2 for a formal definition).

Let μ be a CRM on \mathbb{R}^+ such that $\mathbb{P}[\lim_{t \rightarrow \infty} \mu((0, t]) = \infty] = 1$ and in equation (1) set $Y(t) = \mu((0, t])$. The resulting survival function is

$$S(t) = \mathbb{P}[T > t | \mu] = e^{-\mu(0, t]}, \quad (2)$$

where $\mu(0, t] = \mu((0, t])$. We use the notation $T | \mu \sim \text{NTR}(\mu)$ to denote a random variable with survival function defined as above.

The model defined in equation (2) can be used, for example, to describe the time to remission of a patient in an hospital; precisely, if T_i is the time until death or remission of the i -th patient, the following model

$$T_i | \mu \stackrel{\text{ind.}}{\sim} \text{NTR}(\mu),$$

due to de Finetti's theorem [2], implies that the collection of times to death $\{T_i\}_{i=1}^\infty$ is *exchangeable*, namely the probability law is invariant under permutations that act on a finite number of the collection's indices. This is a desirable property for a survival analysis model. Indeed, we would like the time to death of patients in the same hospital with the same treatment to be exchangeable. Nonetheless, as we have previously highlighted, this model is not appropriate when we consider patients across different hospitals. In this case, it would be desirable to consider a model where patients are exchangeable inside the same hospital but not across hospitals, while at the same time we would like to account for dependence across hospitals.

[10] proposed a natural extension of the model of [6] to deal with these situations. In particular, they propose a bivariate extension by using vectors of completely random measures. We extend their approach to arbitrary dimension by considering d -variate vectors of CRM's $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)$ and corresponding multivariate survival functions given by

$$S(\mathbf{t}) = \mathbb{P}[T_1 > t_1, \dots, T_d > t_d | \boldsymbol{\mu}] = e^{-\mu(0, t_1] - \dots - \mu(0, t_d]}, \quad (3)$$

where $\mathbf{t} = (t_1, \dots, t_d)$. This model is useful when we want to analyse the survival of quantities associated to multiple groups. By considering a d -variate vector of CRM's $\boldsymbol{\mu}$ and d collections of random variables $\{\{T_i^{(j)}\}_{i=1}^\infty\}_{j=1}^d$ such that

$$T_i^{(j)} | \boldsymbol{\mu} \stackrel{\text{ind.}}{\sim} \text{NTR}(\mu_j),$$

we have that each collection $\{T_i^{(j)}\}_{i=1}^\infty$, $j = 1, \dots, d$, is exchangeable; though given $\boldsymbol{\mu}$, any two times to death from different groups $T_l^{(k)}$ and $T_n^{(m)}$ are independent but not identically distributed, so they are not exchangeable. In such context, we say that the collection of times $\{\{T_i^{(j)}\}_{i=1}^\infty\}_{j=1}^d$ is *partially exchangeable*. In this paper, we extend some of the results in [10] to the multivariate setting. The derivation of these results is non-trivial when a dimension greater than two is considered. In particular, we first derive a general expression of the Laplace exponent when a dependence structure induced by

a Lévy copula is considered. We use this result in order to show an explicit expression of the Survival function in terms of the Laplace exponent. We provide a characterization of a d-variate random distribution function in terms of vectors of completely random measures and we fully characterize the posterior vector of completely random measures given the data. From these results, we get two corollaries which are useful for the implementation of the inference.

The paper is organized as follows. Section 2 introduces the concepts of completely random measures, vectors of completely random measures and Lévy copulas. Section 3 introduces the general model and some general results are proven. Section 4 is devoted to the numerical illustration of the model and Section 5 concludes.

2 Preliminaries

In this section, we provide some preliminaries about vectors of completely random measures which are the building block of our Bayesian nonparametric proposal. Furthermore, we will illustrate the concept of a positive Lévy copula which is useful to model the dependence structure between the components of vectors of completely random measures.

2.1 Vectors of completely random measures

The key concept for the construction of the survival analysis models proposed in this work, is the one of completely random measures which we briefly review in this section.

Given a complete and separable metric space \mathbb{X} , with corresponding Borel σ -algebra $\mathcal{X} = \mathcal{B}(\mathbb{X})$, we call a measure μ on $(\mathbb{X}, \mathcal{X})$ boundedly finite if $\mu(A) < \infty$ for any bounded set $A \in \mathcal{X}$. A random measure is a measurable function from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ onto $(\mathbb{M}_{\mathbb{X}}, \mathcal{M}_{\mathbb{X}})$ which is the measure space formed by $\mathbb{M}_{\mathbb{X}}$, the space of boundedly finite measures on $(\mathbb{X}, \mathcal{X})$, and its corresponding Borel σ -algebra $\mathcal{M}_{\mathbb{X}}$. In particular we will focus on the class of *completely random measures* as introduced in [18].

Definition 1. A random measure μ on a complete and separable metric space \mathbb{X} with corresponding Borel σ -algebra $\mathcal{X} = \mathcal{B}(\mathbb{X})$ is called a completely random measure (CRM) if for any collection of disjoint sets $\{A_1, \dots, A_n\} \subset \mathcal{X}$ the random variables $\mu(A_1), \dots, \mu(A_n)$ are mutually independent.

A CRM μ has the following representation [18],

$$\mu = \mu_d + \mu_r + \mu_{fl},$$

where μ_d is a deterministic measure, μ_{fl} is a measure that consists on jumps with possibly random jump heights but fixed jump locations, and

$$\mu_r = \sum_{i=1}^{\infty} W_i \delta_{X_i},$$

where for $i \in \{1, 2, \dots\}$ $X_i \in \mathbb{X}$ are random jump locations and $W_i \in \mathbb{R}^+$ are random jump heights. The measures μ_d , μ_{fl} and μ_r are mutually independent. In particular, μ_r is again a CRM and is characterized by the following Laplace transform

$$\mathbb{E} \left[e^{-\lambda \mu_r(A)} \right] = e^{-\int_{\mathbb{R}^+ \times A} (1 - e^{-\lambda s}) \nu(ds, dx)}, \quad (4)$$

where $\lambda > 0$ and ν is a measure on $\mathbb{R}^+ \times \mathbb{X}$ such that

$$\int_{\mathbb{R}^+ \times A} \min\{s, 1\} \nu(ds, dx) < \infty,$$

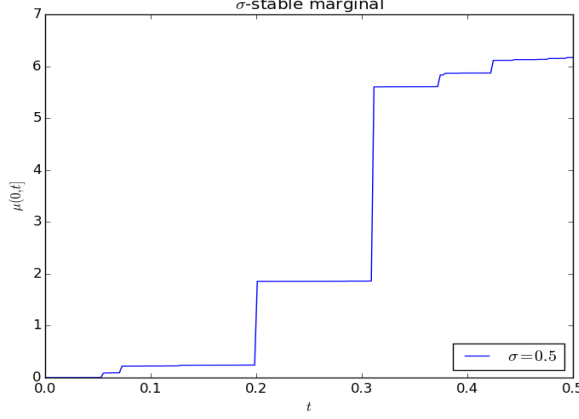


Figure 1: Plot of $Y(t) = \mu(0, t]$ when a σ -stable process is considered.

for any bounded set $A \in \mathcal{X}$. The measure ν is usually called the Lévy intensity of μ_r . In the remainder of this work we only consider CRM's μ without fixed jump locations nor deterministic part so we take $\mu = \mu_r$ to be solely determined by (4). In particular we focus on Lévy intensities ν which are *homogeneous*, i.e.

$$\nu(ds, dx) = \rho(ds)\alpha(dx),$$

where α is a non-atomic measure on \mathbb{X} referring to the jump locations and ρ is a measure on \mathbb{R}^+ referring to the jump heights. A popular example of an homogeneous CRM is the σ -stable process given by

$$\nu(ds, dx) = \frac{A\sigma s^{-1-\sigma}}{\Gamma(1-\sigma)} ds \alpha(dx). \quad (5)$$

As an illustration, we plot in Figure 1 the associated process $Y(t) = \mu(0, t]$ for the σ -stable process (5) with $\alpha(dx) = dx$.

We extend this framework to the multivariate setting by considering vectors (μ_1, \dots, μ_d) where each μ_i is a homogeneous CRM on $(\mathbb{X}, \mathcal{X})$ with respective Lévy intensities $\bar{\nu}_i(ds, dx) = \nu_i(ds)\alpha(dx)$. Moreover we take the intensity α to be additive in the sense that $\alpha((0, t]) = \gamma(t)$ with $\gamma: [0, \infty) \rightarrow \mathbb{R}^+$ a non-decreasing and differentiable function such that $\gamma(0) = 0$ and $\lim_{t \rightarrow \infty} \gamma(t) = \infty$; this last conditions on the limit behaviour will enable us to get, marginally, the associated NTR cumulative distributions in our models.

We have that for any $A, B \in \mathcal{X}$ such that $A \cap B = \emptyset$ the random vectors $(\mu_1(A), \dots, \mu_d(A)), (\mu_1(B), \dots, \mu_d(B))$ are independent; and also one has the multivariate analogue of the Laplace transform (4)

$$\mathbb{E}\left[e^{-\lambda_1 \mu_1(A) - \dots - \lambda_d \mu_d(A)}\right] = e^{-\int_{(\mathbb{R}^+)^d \times A} (1 - e^{-\lambda_1 s_1 - \dots - \lambda_d s_d}) \rho_d(ds_1, \dots, ds_d) \alpha(dx)}, \quad (6)$$

where $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_d) \in (\mathbb{R}^+)^d$ and ρ_d is a measure on $(\mathbb{R}^+)^d$.

In particular we introduce the notation for the multivariate Laplace transform

$$\mathbb{E}\left[e^{-\lambda_1 \mu_1(0, t] - \dots - \lambda_d \mu_d(0, t]}\right] = e^{-\psi_t(\boldsymbol{\lambda})}.$$

Henceforth, $\psi_t(\boldsymbol{\lambda})$ is called the Laplace exponent of $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)$; in the case at hand $\psi_t(\boldsymbol{\lambda}) = \gamma(t)\psi(\boldsymbol{\lambda})$ where $\psi(\boldsymbol{\lambda}) = \int_{(\mathbb{R}^+)^d} (1 - e^{-\langle \boldsymbol{\lambda}, \mathbf{s} \rangle}) \rho_d(d\mathbf{s})$ and $\langle \boldsymbol{\lambda}, \mathbf{s} \rangle = \sum_{i=1}^d \lambda_i s_i$ is the usual inner product in \mathbb{R}^d . Marginalizing, we have that

$$\nu_i(A) = \int_A \nu_i(ds) = \int_{(\mathbb{R}^+)^{d-1}} \rho_d(ds_1, \dots, ds_{i-1}, A, ds_{i+1}, \dots, ds_d).$$

In section 3, we use this particular kind of homogeneous and additive vector of CRM's to construct priors for survival analysis models.

2.2 Positive Lévy copulas

Although in this work we consider vectors of CRM's with fixed marginal behaviour, it remains to establish the dependence structure. [16] introduced the concept of Lévy copulas which allows to construct vectors of Lévy processes with fixed marginals.

Definition 2. A function $\mathcal{C}(\mathbf{s} = (s_1, \dots, s_d)) : [0, \infty)^d \rightarrow [0, \infty]$ is a positive Lévy copula if

- (i) $\forall B = [s_1, t_1] \times \dots \times [s_d, t_d] \subset [0, \infty)^d$ such that $s_1 \leq t_1, \dots, s_d \leq t_d$ we have that

$$\sum_{\{\mathbf{v} : \mathbf{v} \text{ is a vertex of } B\}} \text{sign}(\mathbf{v}) \mathcal{C}(\mathbf{v}) \geq 0,$$

with

$$\text{sign}(\mathbf{v}) = \begin{cases} 1, & \text{if } v_k = s_k \text{ for an even number of vertices,} \\ -1, & \text{if } v_k = s_k \text{ for an odd number of vertices.} \end{cases}$$

- (ii) If \mathbf{s} is such that $s_i = 0$ for some $i \in \{1, \dots, d\}$ then $\mathcal{C}(\mathbf{s}) = 0$.

- (iii) Let $y_1 = \dots = y_{k-1} = y_{k+1} = \dots = y_d = \infty$ and $C_k(s) = \mathcal{C}(y_1, \dots, y_{k-1}, s_k, y_{k+1}, \dots, y_d)$ for $k \in \{1, \dots, d\}$ then $C_k(s) = s$.

For example, a vector of independent Lévy processes is obtained with

$$\mathcal{C}_{\perp, d}(\mathbf{s}) = s_1 \mathbf{1}_{s_2=\infty, \dots, s_d=\infty} + \dots + s_d \mathbf{1}_{s_1=\infty, \dots, s_{d-1}=\infty}.$$

A vector of completely dependent Lévy processes, in the sense that the jumps of the stochastic vector are in a set S such that whenever $\mathbf{v}, \mathbf{u} \in S$ then either $v_i < u_i$ or $u_i < v_i$ for all $i \in \{1, \dots, d\}$, is obtained with

$$\mathcal{C}_{\parallel, d}(\mathbf{s}) = \min\{s_1, \dots, s_d\}.$$

An interesting example of positive Lévy copulas is the Clayton Lévy copula

$$\mathcal{C}_{\theta, d}(\mathbf{s}) = (s_1^{-\theta} + \dots + s_d^{-\theta})^{-\frac{1}{\theta}}, \quad (7)$$

which attains an intermediate level of dependence in the sense that

$$\lim_{\theta \rightarrow 0} \mathcal{C}_{\theta, d}(\mathbf{s}) = \mathcal{C}_{\perp, d}(\mathbf{s}) \text{ and } \lim_{\theta \rightarrow \infty} \mathcal{C}_{\theta, d}(\mathbf{s}) = \mathcal{C}_{\parallel, d}(\mathbf{s}).$$

We define the tail integral of an univariate Lévy intensity ν to be $U(x) = \int_x^\infty \nu(s) ds$. In the setting of section 2.1 we use a Lévy copula \mathcal{C}_d and the marginal tail integrals U_1, \dots, U_d associated to ν_1, \dots, ν_d to specify an absolutely continuous $\rho_d(d\mathbf{s}) = \rho_d(\mathbf{s}) d\mathbf{s}$ via

$$\begin{aligned} U(\mathbf{x}) &= \int_{x_1}^\infty \dots \int_{x_d}^\infty \rho_d(\mathbf{s}) d\mathbf{s} \\ &= \int_{x_1}^\infty \dots \int_{x_d}^\infty \frac{\partial^d}{\partial u_1 \dots \partial u_d} \mathcal{C}_d(\mathbf{u}) \Big|_{u_1=U_1(s_1), \dots, u_d=U_d(s_d)} \nu_1(s_1) \dots \nu_d(s_d) d\mathbf{s}. \end{aligned}$$

So under appropriate regularity conditions we can recover the multivariate Lévy intensity from the copula and marginal intensities in the following way

$$\rho_d(\mathbf{s}) = \frac{\partial^d}{\partial u_1 \dots \partial u_d} \mathcal{C}_d(\mathbf{u}) \Big|_{u_1=U_1(s_1), \dots, u_d=U_d(s_d)} \nu_1(s_1) \dots \nu_d(s_d). \quad (8)$$

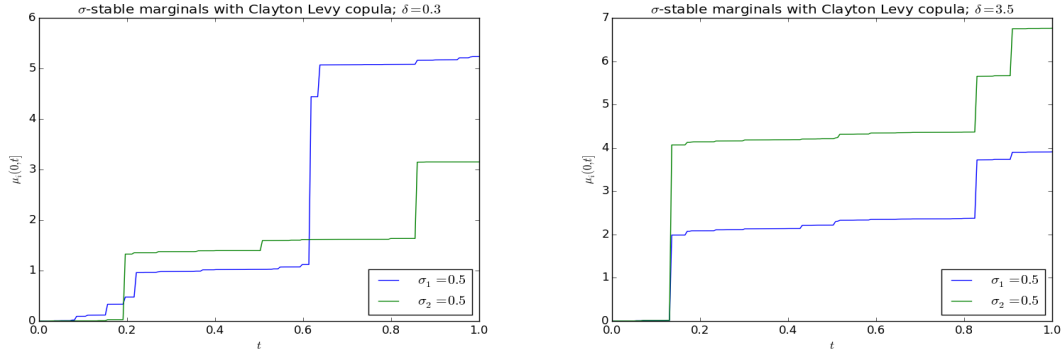


Figure 2: Plot of dependent σ -stable processes with dependence given by Clayton Lévy copula with parameter $\delta = 0.3$ (left) and $\delta = 3.5$ (right).

For example, consider the Clayton Lévy copula with σ -stable margins, given by (5), and $\alpha(dx) = dx$. Figure 2 shows the dependence behaviour when a Clayton Lévy copula with parameter $\delta = 0.3$ and $\delta = 3.5$ is employed. As expected, when $\delta = 0.3$ the processes tend not to jump at the same time, since we are close to the independence case. On the other hand, when δ is increased to 3.5, we can appreciate the higher dependence induced by a larger value of the copula parameter.

2.2.1 Working example

If we consider the Lévy intensity arising from (8) when considering the d -dimensional Clayton Lévy copula, (7), with parameter θ and σ -stable marginals, (5), with parameters A, σ , we obtain

$$\rho_{d,\theta,A,\sigma}(\mathbf{s}) = \frac{A(1+\theta)(1+2\theta)\cdots(1+(d-1)\theta)\sigma^d (s_1 s_2 \cdots s_d)^{\sigma\theta-1}}{\Gamma(1-\sigma) (s_1^{\sigma\theta} + \cdots + s_d^{\sigma\theta})^{\frac{1}{\theta}+d}}.$$

Furthermore, if we take $\theta = 1/\sigma$ we obtain the simplified Lévy intensity

$$\rho_{d,A,\sigma}(\mathbf{s}) = \frac{A(\sigma+1)(\sigma+2)\cdots(\sigma+d-1)\sigma}{\Gamma(1-\sigma) (s_1 + \cdots + s_d)^{\sigma+d}}. \quad (9)$$

Such intensity corresponds to a particular family of vectors of completely random measures known as *compound random measures* and introduced in [12]; the previous Lévy intensity arises when taking $\phi = 1$ in equation (4.4) of the aforementioned paper. A convenient feature of this Lévy intensity is that, as shown in Proposition 3.1 of [29], we can explicitly get the corresponding Laplace exponent

$$\psi_{d,A,\sigma}(\boldsymbol{\lambda}) = \sum_{i=1}^d \frac{\lambda_i^{\sigma+d-1}}{\prod_{j=1, j \neq i}^d (\lambda_i - \lambda_j)}; \quad \lambda_i \neq \lambda_j \text{ for } j \neq i, \quad (10)$$

where we take the appropriate limits when $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_d)$ is such that $\lambda_i = \lambda_j$ for distinct $i, j \in \{1, \dots, d\}$. As indicated in the remark at the end of section 3, evaluation of the Laplace exponent is necessary for the explicit calculation of the posterior mean of the survival function given censored data.

3 Main results

Let $d \in \mathbb{N} \setminus \{0\}$, and suppose we have d collections of random variables

$$\{\{Y_j^{(i)}\}_{j=1}^{\infty}\}_{i=1}^d. \quad (11)$$

We characterize the probability distribution of these random variables in terms of a vector of CRM's $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)$. For $\mathbf{t} = (t_1, \dots, t_d) \in (\mathbb{R}^+)^d$, let

$$S(\mathbf{u}) = \mathbb{P}\left[Y_{i_1}^{(1)} > t_1, \dots, Y_{i_d}^{(d)} > t_d \mid (\mu_1, \dots, \mu_d)\right] = e^{-\mu_1(0, t_1] - \dots - \mu_d(0, t_d]}, \quad (12)$$

where $i_j \in \mathbb{N} \setminus \{0\}$, $j \in \{1, \dots, d\}$. We observe that under such model the random variables (12) are partially exchangeable and marginally follow a NTR process. The dependence structure in this model can be given through the Lévy copula associated to the CRM $\boldsymbol{\mu}$. This model extends the one in [10] to an arbitrary dimension d .

The family of Clayton Lévy copulas is of interest because it has both the independence and complete dependence cases as limit behaviour. In the next result we work towards finding expressions for the Laplace exponent associated to the Clayton family in such a way that the dependence structure is decoupled across dimensions. This result can be useful since, as we will see, an explicit calculation of ψ is of key importance to implement the Bayesian inference in our survival analysis model.

Let $\rho_d(\mathbf{s}; \theta)$ be the Lévy intensity associated via (8) to the Clayton Lévy copula $\mathcal{C}_{\theta, d}$ and fixed marginal Lévy intensities ν_1, \dots, ν_d with corresponding Laplace transforms ψ_1, \dots, ψ_d . We denote the vector of tail integrals corresponding to the marginal Lévy intensities as $\mathbf{U}_d(\mathbf{x}) = (U_1(x_1), \dots, U_d(x_d))$ and fix the notation

$$\kappa(\theta; \boldsymbol{\lambda}, \mathbf{i}) = \lambda_{i_1} \cdots \lambda_{i_d} \int_{(\mathbb{R}^+)^m} \mathbf{e}^{-\lambda_{i_1} s_1 - \dots - \lambda_{i_m} s_m} \mathcal{C}_{\theta, m}(U_{i_1}(s_1), \dots, U_{i_m}(s_m)) \mathbf{d}\mathbf{s},$$

where $d \in \mathbb{N} \setminus \{0\}$, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_d) \in (\mathbb{R}^+)^d$, $m \in \{1, \dots, d\}$, and $\mathbf{i} \in \{1, \dots, d\}^m$ is such that $i_1 < \dots < i_m$.

Proposition 1. Suppose that $d \in \{2, 3, \dots\}$ and

$$\int_{\|\mathbf{s}\| \leq 1} \|\mathbf{s}\| \rho_d(\mathbf{s}; \theta) \mathbf{d}\mathbf{s} < \infty, \quad (13)$$

then

$$\begin{aligned} \psi(\boldsymbol{\lambda}) &= \int_{(\mathbb{R}^+)^d} (1 - \mathbf{e}^{-\langle \boldsymbol{\lambda}, \mathbf{s} \rangle}) \frac{\partial^d}{\partial u_d \cdots \partial u_1} \mathcal{C}_{\theta, m}(\mathbf{u}) \big|_{\mathbf{u}=\mathbf{U}_d(\mathbf{s})} \nu_1(s_1) \cdots \nu_d(s_d) \mathbf{d}\mathbf{s} \\ &= \sum_{i=1}^d \psi_i(\lambda_i) - \sum_{\substack{\mathbf{i}=(i_1, i_2) \in \{1, \dots, d\}^2 \\ i_1 < i_2}} \kappa(\theta; \boldsymbol{\lambda}, \mathbf{i}) + \dots \\ &\quad \dots + (-1)^d \sum_{\substack{\mathbf{i}=(i_1, \dots, i_{d-1}) \in \{1, \dots, d\}^{d-1} \\ i_1 < \dots < i_{d-1}}} \kappa(\theta; \boldsymbol{\lambda}, \mathbf{i}) + (-1)^{d+1} \kappa_d(\theta; \boldsymbol{\lambda}, (1, \dots, d)), \end{aligned}$$

where $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_d) \in (\mathbb{R}^+)^d$.

We refer to the Appendix A.1 for the proof. We incorporate the Lévy exponent ψ in the multivariate survival analysis setting of (12), in the next result. We introduce the notation

$$\nu_{i_1, \dots, i_h}(s_{i_1}, \dots, s_{i_h}) = \int_0^\infty \cdots \int_0^\infty \nu(\mathbf{s}) \prod_{j \notin \{i_1, \dots, i_h\}} \mathbf{d}s_j$$

for $h \in \{1, \dots, d\}$ and distinct $i_1, \dots, i_h \in \{1, \dots, d\}$; and denote ψ_{i_1, \dots, i_h} for the respective Laplace exponents.

Proposition 2. In the context of (12), let $\mathbf{1} = (1, \dots, 1)$. For $t_1 \leq \dots \leq t_d$ and $i_1, \dots, i_d \in \{1, \dots, d\}$ such that $t_{i_1} \leq \dots \leq t_{i_d}$ then

$$\begin{aligned} \mathbb{P}\left[Y^{(1)} > t_1, \dots, Y^{(d)} > t_d\right] &= \\ &\mathbf{e}^{-\gamma(t_{i_1})\psi(\mathbf{1})} \mathbf{e}^{-[\gamma(t_{i_2}) - \gamma(t_{i_1})]\psi_{i_2, \dots, i_d}(\mathbf{1})} \dots \mathbf{e}^{-[\gamma(t_{i_d}) - \gamma(t_{i_{d-1}})]\psi_{i_d}(\mathbf{1})}. \end{aligned} \quad (14)$$

We refer to the Appendix A.2 for the proof. This result showcases the importance of the Laplace exponent ψ for calculating probabilities in the model and the impact of the function $\gamma(t)$, related to the time depending part of the Laplace exponent, in the survival function. In Section 4, we will show that the availability of the Laplace exponent is also of main importance to implement the Bayesian inference for the model. The model we are working on generalizes to arbitrary dimension the classic model of [6]. We present a multivariate extension of Theorem 3.1 in [6], which relates our model with the notion of neutrality to the right. Let F be a d -variate random distribution function on $(\mathbb{R}^+)^d$ and for a d -variate vector of CRM's $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)$ denote $\mu_i(t) = \mu_i((0, t])$ with $i \in \{1, \dots, d\}$. Then we have the next multivariate extension to Theorem 3.1 in [6] and Proposition 4 in [10].

Proposition 3. $F(\mathbf{t} = (t_1, \dots, t_d))$ has the same distribution as

$$[1 - \mathbf{e}^{-\mu_1(t_1)}] \dots [1 - \mathbf{e}^{-\mu_d(t_d)}]$$

for some d -variate CRM $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)$ if and only if for $h \in \{1, 2, \dots\}$ and vectors $\mathbf{t}_1 = (t_{1,1}, \dots, t_{d,1}), \dots, \mathbf{t}_h = (t_{1,h}, \dots, t_{d,h})$ with $t_{0,i} = 0 < t_{1,i} < \dots < t_{d,i}$ and $t_{j,0} = 0 < t_{j,1} < \dots < t_{j,h}$, there exists h independent random vectors $(V_{1,1}, \dots, V_{d,1}), \dots, (V_{1,h}, \dots, V_{d,h})$ such that

$$(F(\mathbf{t}_1), \dots, F(\mathbf{t}_h)) \stackrel{d}{=} \left(V_{1,1} \dots V_{d,1}, [1 - \bar{V}_{1,1} \bar{V}_{1,2}] \dots [1 - \bar{V}_{d,1} \bar{V}_{d,2}], \dots, [1 - \prod_{j=1}^h \bar{V}_{1,j}] \dots [1 - \prod_{j=1}^h \bar{V}_{d,j}] \right), \quad (15)$$

where $\bar{V}_{i,j} = 1 - V_{i,j}$ with $i \in \{1, \dots, d\}$ and $j \in \{1, \dots, h\}$.

We refer to the Appendix A.3 for the proof. We now establish some notation in order to address the posterior distribution arising from (12) when some survival data is available. Let $\mathbf{Y}_{n_i}^{(i)} = (Y_1^{(i)}, \dots, Y_{n_i}^{(i)})$, $i = 1, \dots, d$, be d groups of observations that come from the distribution given by

$$\mathbb{P}[\mathbf{Y}_{n_1}^{(1)} > \mathbf{t}_{1,n_1}, \dots, \mathbf{Y}_{n_d}^{(d)} > \mathbf{t}_{d,n_d} | (\mu_1, \dots, \mu_d)] = \prod_{i=1}^d \prod_{j=1}^{n_i} \mathbf{e}^{-\mu_i(0, t_{i,j}]},$$

where $\mathbf{t}_{i,n_i} = (t_{i,1}, \dots, t_{i,n_i})$ and the event $\{\mathbf{Y}_{n_i}^{(i)} > \mathbf{t}_{i,n_i}\}$ corresponds to the event $\{Y_1^{(i)} > t_{i,1}, \dots, Y_{n_i}^{(i)} > t_{i,n_i}\}$. Let $c_1^{(1)}, \dots, c_{n_1}^{(1)}, \dots, c_1^{(d)}, \dots, c_{n_d}^{(d)}$ be their respective censoring times; so by censored data we mean

$$\mathbf{D} = \bigcup_{i=1}^d \{(T_j^{(i)}, \delta_j^{(i)})\}_{j=1}^{n_i},$$

where $T_j^{(i)} = \min\{Y_j^{(i)}, c_j^{(i)}\}$ and $\delta_j^{(i)} = \mathbb{1}_{(0, c_j^{(i)}]}(Y_j^{(i)})$. The number of exact observations is $n_e = \sum_{i=1}^d \sum_{j=1}^{n_i} \delta_j^{(i)}$ and the number of censored observations is $n_c = n_1 + n_2 - n_e$. Taking into account the possible repetition of values among the observations, we consider the order statistics $(T_{(1)}, \dots, T_{(k)})$ of the distinct observations where k is the number of distinct observed times among all groups.

Let define the set functions

$$m_i^e(A) = \sum_{j=1}^{n_i} \delta_j^{(i)} \mathbb{1}_A(T_j^{(i)}) \quad ; \quad m_i^c(A) = \sum_{j=1}^{n_i} (1 - \delta_j^{(i)}) \mathbb{1}_A(T_j^{(i)})$$

for $i \in \{1, \dots, d\}$, which denote the number of, respectively, exact and censored marginal observations in A , with respect to group i . We define $N_i^e(x) = m_i^e((x, \infty))$, $N_i^c(x) = m_i^c((x, \infty))$, for $i \in \{1, \dots, d\}$ and $n_{i,j}^e = m_i^e(\{T_{(j)}\})$, $n_{i,j}^c = m_i^c(\{T_{(j)}\})$, $\bar{n}_{i,j}^e = \sum_{r=j}^k n_{i,r}^e$, $\bar{n}_{i,j}^c = \sum_{r=j}^k n_{i,r}^c$ for $(i, j) \in$

$\{1, \dots, d\} \times \{1, \dots, k\}$; and the corresponding vectors $\bar{\mathbf{n}}_j^e = (\bar{n}_{1,j}^e, \dots, \bar{n}_{d,j}^e)$, $\bar{\mathbf{n}}_j^c = (\bar{n}_{1,j}^c, \dots, \bar{n}_{d,j}^c)$, for $j \in \{1, \dots, k\}$ and $\mathbf{N}^e(x) = (N_1^e(x), \dots, N_d^e(x))$, $\mathbf{N}^c(x) = (N_1^c(x), \dots, N_d^c(x))$. The next theorem determines the calculation of the posterior distribution for a vector of CRM's given some censored data, it applies to general vectors of CRM's, in particular the assumption that the respective Lèvy intensity is homogeneous has been dropped.

Theorem 1. Let $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)$ be a d -variate CRM such that its corresponding Lèvy intensity $\nu(\mathbf{s}, d\mathbf{t})d\mathbf{s}$ is differentiable with respect to t_0 on $\mathbb{R}^+ \setminus \{0\}$ in the sense that for $\eta_t = \nu(\mathbf{s}, (0, t])$ the partial derivative $\eta'_{t_0}(\mathbf{s}) = \partial \eta_t(\mathbf{s}) / \partial t|_{t=t_0}$ exists. Moreover we assume that the entries of $\boldsymbol{\mu}$ are not independent. Then the posterior distribution of $\boldsymbol{\mu}$ given data \mathbf{D} is the distribution of the random measure

$$(\mu_1^*, \dots, \mu_d^*) + \sum_{\{j : T_{(j)} \text{ is an exact observation}\}} (J_{1,j} \delta_{T_{(j)}} + \dots + J_{d,j} \delta_{T_{(j)}})$$

where

i) $\boldsymbol{\mu}^* = (\mu_1^*, \dots, \mu_d^*)$ is a d -variate CRM with Lèvy intensity ν^* such that

$$\nu(d\mathbf{s}, d\mathbf{x})|_{d\mathbf{x} \in (T_{(j-1)}, T_{(j)})} = \mathbf{e}^{-\langle \bar{\mathbf{n}}_j^c + \bar{\mathbf{n}}_j^e, \mathbf{s} \rangle} \nu(d\mathbf{s}, d\mathbf{x})$$

for $j \in \{1, \dots, k+1\}$.

ii) The vectors of jumps $\{(J_{1,j}, \dots, J_{d,j})\}_{j \in J}$, with $J = \{j : T_{(j)} \text{ is an exact observation}\}$, are mutually independent and the vector of jumps corresponding to the exact observation $T_{(j)}$ has density

$$f_j(\mathbf{s}) \propto \prod_{i=1}^d \left\{ \mathbf{e}^{-(\bar{n}_{i,j}^c + \bar{n}_{i,j+1}^e)s_i} (1 - \mathbf{e}^{-s_i})^{n_{i,j}^e} \right\} \eta'_{T_{(j)}}(\mathbf{s}).$$

iii) The random measure $\boldsymbol{\mu}^*$ is independent of $\{(J_{1,j}, \dots, J_{d,j})\}_{j \in J}$, with $J = \{j : T_{(j)} \text{ is an exact observation}\}$.

We refer to the Appendix A.4 for the proof. The previous result showcases that the posterior distribution arising from (12) can be modeled in the same framework via a vector of CRM's by updating the prior vector of CRM's $\boldsymbol{\mu}$ to $\boldsymbol{\mu}^*$ as indicated in the theorem.

This result is enough to give a scheme for the inference in the model. In particular, in the setting of (12) and Theorem 1, we want to estimate the corresponding multivariate survival function $\mathbb{P}[Y^{(1)} > t_1, \dots, Y^{(d)} > t_d | (\mu_1, \dots, \mu_d)]$.

Given censored data \mathbf{D} , the natural Bayesian estimator would be the posterior mean of the survival function given the data. As a result of Theorem 1 we can calculate such quantity. The next allows us to implement the necessary inferential scheme for performing the estimation of the survival function as a posterior mean. We denote \mathbf{e}_i for the canonical basis of \mathbb{R}^d , and $S_L(t) = S(t \sum_{l \in L} \mathbf{e}_l)$ for $t > 0$, $\emptyset \neq L \subset \{1, \dots, d\}$. In view of the independent increments of the CRM's, calculation of the posterior mean of S_L is all that is needed for evaluation of the posterior mean of S . The next corollary shows how to evaluate the posterior mean of S_L .

Corollary 1. Let $\boldsymbol{\mu}$ be a vector of CRM's with corresponding Lèvy intensity such that $\eta_t(\mathbf{s}) = \gamma(t)\nu(\mathbf{s})$ with γ a differentiable function satisfying $\gamma'(t) \neq 0$ for $t > 0$. Moreover we assume that the entries of $\boldsymbol{\mu}$ are not independent. Let $\emptyset \neq L \subset \{1, \dots, d\}$ and set

$$J_t = \{j : T_{(j)} \leq t\}$$

where we let $T_{(k+1)} = \infty$. Then

$$\begin{aligned} \hat{S}_L(t) &= \mathbb{E}[\mathbb{E}[S_L(t)|\boldsymbol{\mu}] | \mathbf{D}] = \mathbf{e}^{-\sum_{j=1}^{k+1} [\gamma(t \wedge T_{(j)}) - \gamma(T_{(j-1)})] \mathbf{1}_{[T_{(j-1)}, \infty)}(t)} \psi_j^*(\sum_{l \in L} \mathbf{e}_l) \\ &\times \prod_{j \in J_t} \gamma'(T_{(j)}) \left[\frac{\int_{(\mathbb{R}^+)^d} \prod_{i=1}^d \left\{ \mathbf{e}^{-(\mathbf{1}_{i \in L} + \bar{\mathbf{n}}_{i,j}^c + \bar{\mathbf{n}}_{i,j+1}^e) s_i} (1 - \mathbf{e}^{-s_i})^{n_{i,j}^e} \right\} \nu(\mathbf{s}) d\mathbf{s}}{\int_{(\mathbb{R}^+)^d} \prod_{i=1}^d \left\{ \mathbf{e}^{-(\bar{\mathbf{n}}_{i,j}^c + \bar{\mathbf{n}}_{i,j+1}^e) s_i} (1 - \mathbf{e}^{-s_i})^{n_{i,j}^e} \right\} \nu(\mathbf{s}) d\mathbf{s}} \right] \end{aligned}$$

where for $\boldsymbol{\lambda} \in (\mathbb{R}^+)^d$

$$\begin{aligned} \psi_j^*(\boldsymbol{\lambda}) &= \int_{(\mathbb{R}^+)^d} (1 - \mathbf{e}^{-\langle \boldsymbol{\lambda}, \mathbf{s} \rangle}) \mathbf{e}^{-(\bar{\mathbf{n}}_j^c + \bar{\mathbf{n}}_j^e, \mathbf{s})} \nu(\mathbf{s}) d\mathbf{s} \\ &= \psi(\boldsymbol{\lambda} + \bar{\mathbf{n}}_j^c + \bar{\mathbf{n}}_j^e) - \psi(\bar{\mathbf{n}}_j^c + \bar{\mathbf{n}}_j^e). \end{aligned}$$

We see that we can estimate $S(\mathbf{t})$ for arbitrary $\mathbf{t} \in (\mathbb{R}^+)^d$ in terms of the estimates defined in the previous corollary. Indeed, let $\mathbf{t} = (t_1, \dots, t_d)$ and π be a permutation of $\{1, \dots, d\}$ such that $t_{\pi(1)} \leq t_{\pi(2)} \leq \dots \leq t_{\pi(d)}$. We define for $i \in \{1, \dots, d-1\}$ sets

$$L_i = \{j : \sigma^{(-1)}(j) \geq i\}.$$

It follows from the independence of increment of CRM's that for $\mathbf{t} \in (\mathbb{R}^+)^d$ the posterior mean of the survival function given censored data \mathbf{D} is

$$\hat{S}(\mathbf{t}) = \mathbb{E}[\mathbb{E}[S(\mathbf{t})|\boldsymbol{\mu}] | \mathbf{D}] = \hat{S}_{L_1}(t_{\pi(1)}) \prod_{i=1}^{d-1} \frac{\hat{S}_{L_i}(t_{\pi(i+1)})}{\hat{S}_{L_i}(t_{\pi(i)})}. \quad (16)$$

Usually the Lévy intensities we deal with have some dependence in a vector of hyper-parameters \mathbf{c} . On the proof of Theorem 1 it is outlined how to, given censored data \mathbf{D} as before, get the likelihood of the hyper-parameters in the Lévy intensity as the next corollary shows:

Corollary 2. Let $\boldsymbol{\mu}$ be a vector of CRM's with corresponding Lévy intensity such that $\eta_t(\mathbf{s}) = \gamma(t) \rho_{d,\mathbf{c}}(\mathbf{s})$ with γ a differentiable function satisfying $\gamma'(t) \neq 0$ for $t > 0$, and \mathbf{c} a vector of hyper-parameters. Given censored data \mathbf{D} we get the likelihood on \mathbf{c} .

$$\begin{aligned} l(\mathbf{c}; \mathbf{D}) &= \mathbf{e}^{-\sum_{j=1}^k [\gamma(T_{(j)}) - \gamma(T_{(j-1)})] \psi_{d,\mathbf{c}}(\bar{\mathbf{n}}_j^c + \bar{\mathbf{n}}_j^e)} \\ &\times \prod_{j \in J} \gamma'(T_{(j)}) \int_{(\mathbb{R}^+)^d} \prod_{i=1}^d \left\{ \mathbf{e}^{-(\bar{\mathbf{n}}_{i,j}^c + \bar{\mathbf{n}}_{i,j+1}^e) s_i} (1 - \mathbf{e}^{-s_i})^{n_{i,j}^e} \rho_{d,\mathbf{c}}(\mathbf{s}) d\mathbf{s} \right\}, \end{aligned}$$

where $\psi_{d,\mathbf{c}}$ is the Laplace exponent associated to $\rho_{d,\mathbf{c}}$.

The next lemma provides a useful identity for the computation of integrals as the appearing in the equation above and also in Corollary 1.

Lemma 1. For $\mathbf{q} = (q_1, \dots, q_d) \in (\mathbb{R}^+)^d$ and $\mathbf{n} = (n_1, \dots, n_d) \in \mathbb{N}^d$

$$\begin{aligned} \int_{(\mathbb{R}^+)^d} \mathbf{e}^{-\langle \mathbf{q}, \mathbf{x} \rangle} \prod_{i=1}^d (1 - \mathbf{e}^{-s_i})^{n_i} \nu(\mathbf{s}) d\mathbf{s} &= \sum_{i=1}^d \sum_{k=1}^{n_i} \binom{n_i}{k} (-1)^{k+1} [\psi(k\mathbf{e}_i + \mathbf{q}) - \psi(\mathbf{q})] \\ &+ \sum_{\substack{i_1 \neq i_2 \\ n_{i_1}, n_{i_2} \notin \{0\}}} \sum_{k_1=1}^{n_{i_1}} \sum_{k_2=1}^{n_{i_2}} \binom{n_{i_1}}{k_1} \binom{n_{i_2}}{k_2} (-1)^{k_1+k_2+1} [\psi(k_1\mathbf{e}_{i_1} + k_2\mathbf{e}_{i_2} + \mathbf{q}) - \psi(\mathbf{q})] \\ &+ \dots + \mathbf{1}_{\{n_1 \neq 0, \dots, n_d \neq 0\}} \sum_{k_1=1}^{n_1} \dots \sum_{k_d=1}^{n_d} (-1)^{k_1+\dots+k_d+1} [\psi(k_1\mathbf{e}_1 + \dots + k_d\mathbf{e}_d) - \psi(\mathbf{q})]. \end{aligned}$$

We omit the proof as it is just an application of the binomial theorem in the same line as the proof of Lemma 5 in the appendix.

Remark. We see with the previous lemma and corollaries that implementation of the inference problem depends on whether we can perform evaluations of the Laplace exponent or not.

4 Applications

In this section we perform the fitting of a multivariate survival function given censored to the right data in the framework of (12). As mentioned in the previous remark, the evaluation of the Laplace exponent of μ in (12) is necessary to evaluate the posterior mean in Corollary 1 and the likelihood in Corollary 2; with this in mind we choose the random measure μ given by the Lévy intensity showcased in (9), so that the corresponding Laplace exponent is readily given by (10). For illustration purposes we use 4-dimensional data arising from a distributional copula with fixed marginal distributions, see [26] for an overview of distributional copulas.

More precisely, we generate simulated data $\mathbf{Y} = (Y_1, \dots, Y_4)$ with probability distribution $F_{\theta, \lambda}$ given by a distributional Clayton copula with parameter θ and exponential marginals with parameter λ . Then, we perform right-censoring by considering censoring time variables \mathbf{c} consisting of independent exponential random variables with parameter λ_c , and define

$$\begin{aligned} \delta &= (\mathbf{1}_{Y_1 < c_1}, \dots, \mathbf{1}_{Y_4 < c_4}), \\ \mathbf{T} &= (\min\{Y_1, c_1\}, \dots, \min\{Y_4, c_4\}). \end{aligned} \quad (17)$$

For the fit we use the 4-dimensional Lévy intensity given by (9) and assign priors for the hyperparameters in (9), α and A . We choose a log-normal prior for the parameter A and a Beta prior for the parameter α . We use the Metropolis within Gibbs algorithm to draw samples from the posterior distributions of A and α by making use of the likelihood showed in Corollary 2. We present a Monte Carlo approximation of the estimator (16), where we have averaged over the posterior draws of A and α .

In figure 3 we show the fit for 150 observations of the from (17) to different values of $\hat{S}(t_1, \dots, t_4)$ where we have chosen

$$\begin{aligned} \mathbf{Y}_j &\sim F_{\theta=0.3, \lambda=1}, & j &= 1, \dots, 150 \\ c_{i,j} &\sim \text{Exp}(\lambda_c = 3.7), & i &= 1, \dots, 4; \quad j = 1, \dots, 150 \\ T_{i,j} &= \min\{Y_{i,j}, c_{i,j}\}, & i &= 1, \dots, 4; \quad j = 1, \dots, 150. \end{aligned}$$

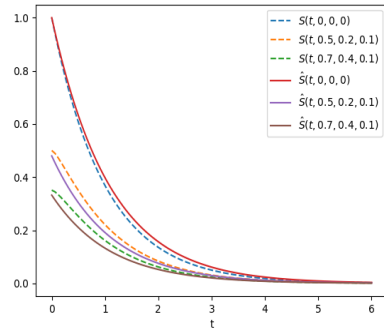
We chose $\lambda_c = 3.7$ so that we have at least 75% of exact observations for \mathbf{T} in each dimension. For the prior distributions in the hyperparameters we selected

$$\begin{aligned} \alpha &\sim \text{Beta}(\mu = 0.4, \sigma^2 = 0.1) \\ A &\sim \text{Log-Norm}(\mu = \log(0.88), \sigma^2 = 3.5). \end{aligned}$$

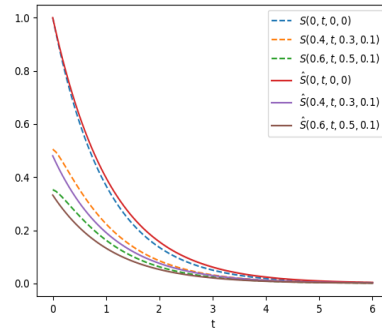
We ran 1000 iterations for the associated Gibbs sampler. We see in Figure 3 that the estimated survival functions approximate well the true functions.

5 Conclusions

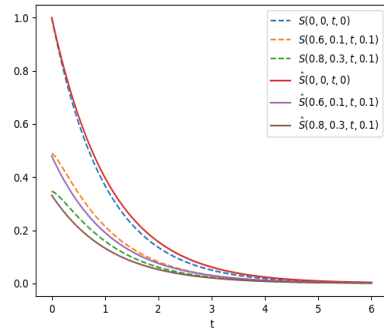
Several specific constructions have been proposed in Bayesian nonparametrics for vectors of completely random measures, including various forms of superposition [13, 22, 23, 3] and Lévy copula-based approaches [19, 20, 29]. In this paper, we focused on the use of vectors of Bayesian nonparametric priors for estimating the Survival function. In particular, we have extended the results of [10] to



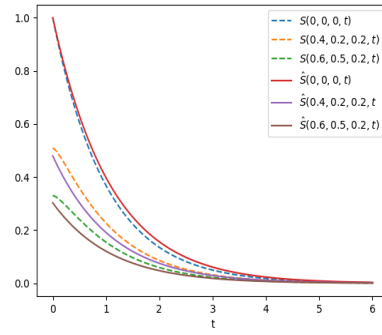
(a) Dimension 1



(b) Dimension 2



(c) Dimension 3



(d) Dimension 4

Figure 3: Plot of the fits for survival functions associated to $F_{\theta=0.3, \lambda=1.}$

dimensions higher than two. Such extension allows for the modelling partially exchangeable data sets, e.g. when the number of groups sharing exchangeable observations is allowed to be any natural number. Further to the use of Levy copulas, [15] used an invariance principle to propose a vector of CRM's, and more recently [12] gave a flexible scheme to construct vectors of CRM's. Such constructions can be used in the context of Theorem 1 to calculate the posterior distribution in the multivariate NTR model considered. For illustration purposes we have showcased a simulation study where the multivariate neutral to the right model approximates the survival function related to a distributional Clayton copula with equal marginal distributions. The estimates given in the application section consist on the calculation of posterior means, which in the present work has relied on the evaluation of the Laplace exponent. In future work we want to allow for inference when an analytical expression of the Laplace exponent is not known, and also for the calculation of credible intervals for the survival functions.

Acknowledgements

Fabrizio Leisen was supported by the European Communitys Seventh Framework Programme under grant agreement no: 630677. The scholarship of Alan Riva Palacio is partially funded by the Marie Curie Career Integration grant Flexible Bayesian nonparametric prior (European Communitys Seventh Framework Programme under grant agreement no: 630677).

Appendix

A.1 Proof of Proposition 1

Given $d \in \{2, 3, \dots\}$ we use the notation $v_{-i}(\mathbf{s}) = \prod_{j=i+1}^d v_j(s_j)$ and $\mathbf{U}_{k:d}(\mathbf{s}) = (U_k(s_1), \dots, U_d(s_{d-k+1}))$ for $\mathbf{s} \in (\mathbb{R}^+)^d$. Furthermore we define integrals

$$a_{0,m}(\boldsymbol{\lambda}) = \int_{(\mathbb{R}^+)^m} (1 - \mathbf{e}^{-\langle \boldsymbol{\lambda}, \mathbf{s} \rangle}) \frac{\partial^d}{\partial u_d \cdots \partial u_1} C_{\theta,m}(\mathbf{u})|_{\mathbf{u}=\mathbf{U}_{d-m+1:d}(\mathbf{s})} v_{-0}(\mathbf{s}) d\mathbf{s}$$

and

$$a_{k,m}(\boldsymbol{\lambda}) = (-1)^{k+1} \int_{(\mathbb{R}^+)^m} \lambda_1 \cdots \lambda_k \mathbf{e}^{-\langle \boldsymbol{\lambda}, \mathbf{s} \rangle} \frac{\partial^{d-k}}{\partial u_d \cdots \partial u_{k+1}} C_{\theta,m}(\mathbf{U}_{d-m+1:d}(\mathbf{s})) v_{-k}(\mathbf{s}) d\mathbf{s}$$

where $k \in \{1, \dots, d\}$, $m \in \{0, 1, \dots, d\}$ and $\boldsymbol{\lambda} \in (\mathbb{R}^+)^d$ such that $a_{0,d}(\boldsymbol{\lambda}) < \infty$; we also define $\prod_{j=k}^l a_j = 1$ when $k > l$, and denote \mathbf{x}_{-i} for the vector \mathbf{x} without its i -th entry.

An integration by parts shows that

$$\begin{aligned} a_{0,d} &= - \int_{(\mathbb{R}^+)^{d-1}} (1 - \mathbf{e}^{-\langle \boldsymbol{\lambda}, \mathbf{s} \rangle}) \frac{\partial^{d-1}}{\partial u_d \cdots \partial u_2} C_{\theta,d}(\mathbf{u})|_{\mathbf{u}=\mathbf{U}_d(\mathbf{s})} v_{-1}(\mathbf{s}) \bigg|_{s_1=0}^{s_1=\infty} d\mathbf{s}_{-1} \\ &\quad + \int_{(\mathbb{R}^+)^d} \lambda_1 \mathbf{e}^{-\langle \boldsymbol{\lambda}, \mathbf{s} \rangle} \frac{\partial^{d-1}}{\partial u_d \cdots \partial u_2} C_{\theta,d}(\mathbf{u})|_{\mathbf{u}=\mathbf{U}_d(\mathbf{s})} v_{-1}(\mathbf{s}) d\mathbf{s} \\ &= a_{0,d-1}(\boldsymbol{\lambda}_{-1}) + a_{1,d}(\boldsymbol{\lambda}) \end{aligned}$$

and in general for $r \in \{1, \dots, d\}$ we get the recursion formula

$$a_{r,d}(\boldsymbol{\lambda}) = a_{r,d-1}(\boldsymbol{\lambda}_{-(r+1)}) + a_{r+1,d}(\boldsymbol{\lambda}) \quad (\text{A.18})$$

We prove the next technical lemma

Lemma 2. If $a_{0,d}(\lambda) < \infty$ then the next $d + 1$ identities hold

$$\begin{aligned}
a_{0,d}(\lambda) &= \sum_{i=1}^d \psi_i(\lambda_i) - \sum_{\substack{i=(i_1,i_2) \in \{1,\dots,d\}^2 \\ i_1 < i_2}} \kappa(\theta; \lambda, i) + \dots \\
&\quad \dots + (-1)^d \sum_{\substack{i=(i_1,\dots,i_{d-1}) \in \{1,\dots,d\}^{d-1} \\ i_1 < \dots < i_{d-1}}} \kappa(\theta; \lambda, (i_1, \dots, i_{d-1})) + (-1)^{d+1} \kappa_d(\theta; \lambda) \\
a_{1,d}(\lambda) &= \psi_1(\lambda_1) - \sum_{i=2}^d \kappa(\theta; \lambda, (1, i)) + \sum_{\substack{i_1, i_2 \in \{2,\dots,d\} \\ i_1 < i_2}} \kappa(\theta; \lambda, (1, i_1, i_2)) + \dots \\
&\quad \dots + (-1)^d \sum_{\substack{i_1, \dots, i_{d-2} \in \{2,\dots,d\} \\ i_1 < \dots < i_{d-2}}} \kappa(\theta; \lambda, (1, i_1, \dots, i_{d-2})) + (-1)^{d+1} \kappa(\theta; \lambda, (1, \dots, d)) \\
&\quad \vdots \\
a_{d-1,d}(\lambda) &= (-1)^d \kappa(\theta; \lambda, (1, \dots, d-1)) + (-1)^{d+1} \kappa(\theta; \lambda, (1, \dots, d)) \\
a_{d,d}(\lambda) &= (-1)^{d+1} \kappa(\theta; \lambda, (1, \dots, d))
\end{aligned} \tag{A.19}$$

Proof. We proceed by mathematical induction over the dimension d . We observe that from the definition of κ we always have

$$a_{d,d}(\lambda) = (-1)^{d+1} \kappa(\theta; \lambda, (1, \dots, d))$$

For the case $d = 2$ we have from Proposition 1 in [10] that

$$a_{0,2}(\lambda_1, \lambda_2) = \psi_1(\lambda_1) + \psi_2(\lambda_2) - \kappa_2(\theta; \lambda_1, \lambda_2)$$

And integrating by parts we obtain

$$\begin{aligned}
a_{1,2}(\lambda_1, \lambda_2) &= \int_{\mathbb{R}^+} \lambda_1 e^{-\lambda_1 s_1} U_1(x_1) ds_1 - \lambda_1 \lambda_2 \int_{(\mathbb{R}^+)^2} e^{-\lambda_1 x_2 - \lambda_2 s_2} C_\theta(U_1(s_1), U_2(s_2)) ds_1 ds_2 \\
&= \psi_1(\lambda_1) - \kappa(\theta; \lambda, (1, 2))
\end{aligned}$$

So we get the validity of the equations in (A.19) for the case $d = 2$. Now we suppose that (A.19) is true for $d = m - 1$, we must show the validity for $d = m$. From the recursion formula (A.18) we get for $r \in \{0, 1, \dots, d\}$

$$a_{r,m}(\lambda) = a_{r,m-1}(\lambda_{-(r+1)}) + a_{r+1,m-1}(\lambda_{-(r+2)}) + \dots + a_{m-1,m-1}(\lambda_{-m}) + a_{m,m}(\lambda)$$

So the validity of (A.19) for $d = m$ follows from the validity for $d = m - 1$ and a combinatorial argument. \square

Proposition 1 follows by considering the first equation in the Lemma statement and the definition of $a_{0,d}$.

A.2 Proof of Proposition 2

Proof. Using the independent increments property of CRM's we get that

$$\begin{aligned}
\mathbb{P}[Y^{(1)} > t_1, \dots, Y^{(d)} > t_d] &= \mathbb{E}[e^{-\mu_1(0,t_1] - \dots - \mu_d(0,t_d)}] \\
&= \mathbb{E}[e^{-\mu_{i_1}(0,t_{i_1}] - \dots - \mu_{i_d}(0,t_{i_d})}] \mathbb{E}[e^{-\mu_{i_2}(t_{i_1},t_{i_2}] - \dots - \mu_{i_d}(t_{i_1},t_{i_2})}] \dots \mathbb{E}[e^{-\mu_{i_d}(t_{i_{d-1}},t_{i_d})}] \\
&= e^{-\gamma(t_{i_1})} \psi(\mathbf{1}) e^{-[\gamma(t_{i_2}) - \gamma(t_{i_1})]} \psi_{i_2, \dots, i_d}(\mathbf{1}) \dots e^{-[\gamma(t_{i_d}) - \gamma(t_{i_{d-1}})]} \psi_{i_d}(\mathbf{1})
\end{aligned}$$

\square

A.3 Proof of Proposition 3

Proof. For the only if part we define $V_{i,j} = 1 - \mathbf{e}^{-[\mu_i(t_{i,j}) - \mu_i(t_{i,j-1})]}$ for $i \in \{1, \dots, d\}$ and $j \in \{1, \dots, h\}$ so by supposing $(F_1(t_1), \dots, F_d(t_d)) \stackrel{d}{=} (1 - \mathbf{e}^{-\mu_1 t_1}, \dots, 1 - \mathbf{e}^{-\mu_d t_d})$ we have

$$\begin{aligned} F(t_{1,1}, \dots, t_{d,1}) &\stackrel{d}{=} [1 - \mathbf{e}^{-\mu_1(t_{1,1})}] \dots [1 - \mathbf{e}^{-\mu_d(t_{d,1})}] \\ &= [1 - \mathbf{e}^{-[\mu_1(t_{1,1}) - \mu_1(t_{1,0})]}] \dots [1 - \mathbf{e}^{-[\mu_d(t_{d,1}) - \mu_d(t_{d,0})]}] \\ &= V_{1,1} \dots V_{d,1} \end{aligned}$$

We observe that for $i \in \{2, \dots, h\}$ and $r \in \{1, \dots, d\}$

$$1 - \prod_{j=1}^i \bar{V}_{r,j} = 1 - \prod_{j=1}^i (1 - V_{r,j}) = 1 - \prod_{j=1}^i \mathbf{e}^{-[\mu_r(t_{r,j}) - \mu_r(t_{r,j-1})]} = 1 - \mathbf{e}^{-\mu_r(t_{r,i})}$$

So for $i \in \{2, \dots, d\}$

$$\begin{aligned} F(t_{1,i}, \dots, t_{d,i}) &\stackrel{d}{=} [1 - \mathbf{e}^{-\mu_1(t_{1,i})}] \dots [1 - \mathbf{e}^{-\mu_d(t_{d,i})}] \\ &= [1 - \prod_{j=1}^i \bar{V}_{1,j}] \dots [1 - \prod_{j=1}^i \bar{V}_{d,j}] \end{aligned}$$

For the if part we define $\mu_i(t) = -\log(1 - F_i(t))$ for $i \in \{1, \dots, d\}$ and suppose for $h \in \{1, 2, \dots\}$, $\mathbf{t}_1 = (t_{1,1}, \dots, t_{d,1}), \dots, \mathbf{t}_h = (t_{1,h}, \dots, t_{d,h})$ with $t_{0,i} = 0 < t_{1,i} < \dots < t_{d,i}$ and $t_{j,0} = 0 < t_{j,1} < \dots < t_{j,h}$ the existence of independent random vectors $(V_{1,1}, \dots, V_{d,1}), \dots, (V_{1,h}, \dots, V_{d,h})$ such that we have (15). Marginalizing in (15), we can apply Theorem 3.1 of [6] to each F_i so we obtain that every marginal process $\mu_{i,t}$ is stochastically continuous, almost surely non-decreasing and has the appropriate limit behaviour.

We observe that

$$(\mu_1(t_j) - \mu_1(t_{j-1}), \dots, \mu_d(t_j) - \mu_d(t_{j-1})) \stackrel{d}{=} (-\log(1 - V_{1,j}), \dots, -\log(1 - V_{d,j}))$$

Hence $(\mu_{1,t}, \dots, \mu_{d,t})$ defines a CRM. □

A.4 Proof of Theorem 1

In order to prove the theorem we use the next technical lemma.

Lemma 3. Let (μ_1, \dots, μ_d) be a d -variate CRM such that μ_1, \dots, μ_d are not independent and let the Lévy intensity $\nu(\mathbf{s}, d\mathbf{s})$ of (μ_1, \dots, μ_d) be such that $\eta_t = \nu(\mathbf{x}, (0, t])$ is differentiable with respect to $t \in \mathbb{R}^+$ at some $t_0 \neq 0$ and denote $\eta'_{t_0}(\mathbf{s}) = \partial \eta_t(\mathbf{s}) / \partial t|_{t=t_0}$. If $\mathbf{q} = (q_1, \dots, q_d) \in \mathbb{N}^d$ are such that $\max\{q_1, \dots, q_d\} \geq 1$ and $\mathbf{r} = (r_1, \dots, r_d) \in (\mathbb{R}^+)^d$ are such that $\min\{r_1, \dots, r_d\} \geq 1$, then

$$\begin{aligned} \mathbb{E} \left[\mathbf{e}^{-r_1 \mu_1(A_\epsilon) - \dots - r_d \mu_d(A_\epsilon)} \left(1 - \mathbf{e}^{-\mu_1(A_\epsilon)} \right)^{q_1} \dots \left(1 - \mathbf{e}^{-\mu_d(A_\epsilon)} \right)^{q_d} \right] \\ = \epsilon \int_{(\mathbb{R}^+)^d} \mathbf{e}^{-\langle \mathbf{r}, \mathbf{s} \rangle} (1 - \mathbf{e}^{-s_1})^{q_1} \dots (1 - \mathbf{e}^{-s_d})^{q_d} \eta'_{t_0}(\mathbf{s}) d\mathbf{s} + o(\epsilon) \end{aligned}$$

as $0 < \epsilon \rightarrow 0$, with $A_\epsilon = (t_0 - \epsilon, t_0]$ for some $t_0 \in \mathbb{R}^+ \setminus \{0\}$.

Proof. We denote $\triangle_{s_1}^{s_2} f_t(\mathbf{r}) = f_{s_2}(\mathbf{r}) - f_{s_1}(\mathbf{r})$ for a function f where $s_1, s_2 \in \mathbb{R}^+$ and $\mathbf{r} \in \mathbb{R}^d$. We use the

binomial theorem and apply expectation to write the left hand side in the equation above as

$$\begin{aligned}
& \sum_{j_1=0}^{q_1} \cdots \sum_{j_d=0}^{q_d} \binom{q_1}{j_1} \cdots \binom{q_d}{j_d} (-1)^{j_1+\cdots+j_d} \mathbf{e}^{-[\psi_{t_0}(r_1+j_1, \dots, r_d+j_d) - \psi_{t_0-\epsilon}(r_1+j_1, \dots, r_d+j_d)]} \\
&= \mathbf{e}^{-\Delta_{t_0-\epsilon}^{t_0} \psi_t(\mathbf{r})} + \mathbf{e}^{-\Delta_{t_0-\epsilon}^{t_0} \psi_t(\mathbf{r})} \left\{ \sum_{i=1}^d \sum_{j=1}^{q_i} \binom{q_i}{j} (-1)^j \mathbf{e}^{-\Delta_{t_0-\epsilon}^{t_0} [\psi_t(\mathbf{r}+j\mathbf{e}_i) - \psi_t(\mathbf{r})]} \right. \\
&\quad + \sum_{\substack{i_1, i_2 \in \{1, \dots, d\} \\ i_1 < i_2}} \sum_{j_1=1}^{q_{i_1}} \sum_{j_2=1}^{q_{i_2}} \binom{q_{i_1}}{j_1} \binom{q_{i_2}}{j_2} (-1)^{j_1+j_2} \mathbf{e}^{-\Delta_{t_0-\epsilon}^{t_0} [\psi_t(\mathbf{r}+j_1\mathbf{e}_{i_1}+j_2\mathbf{e}_{i_2}) - \psi_t(\mathbf{r})]} \\
&\quad \left. + \cdots + \sum_{j_1=1}^{q_1} \cdots \sum_{j_d=1}^{q_d} \binom{q_1}{j_1} \cdots \binom{q_d}{j_d} (-1)^{\langle \mathbf{1}, \mathbf{j} \rangle} \mathbf{e}^{-\Delta_{t_0-\epsilon}^{t_0} [\psi_t(\mathbf{r}+\mathbf{j}) - \psi_t(\mathbf{r})]} \right\} \quad (\text{D.20})
\end{aligned}$$

We note that for $j_i \in \{0, \dots, x_i\}$, $i \in \{1, \dots, d\}$, $\mathbf{j} = (j_1, \dots, j_d)$, a Taylor expansion yields

$$\begin{aligned}
& \mathbf{e}^{-\Delta_{t_0-\epsilon}^{t_0} [\psi_t(\mathbf{r}+\mathbf{j}) - \psi_t(\mathbf{r})]} = \mathbf{e}^{-\int_{(\mathbb{R}^+)^d} \mathbf{e}^{-\langle \mathbf{r}, \mathbf{s} \rangle} (1 - \mathbf{e}^{-\langle \mathbf{j}, \mathbf{s} \rangle}) \Delta_{t_0-\epsilon}^{t_0} \eta_t(\mathbf{s}) d\mathbf{s}} \\
&= 1 - \epsilon \int_{(\mathbb{R}^+)^d} \mathbf{e}^{-\langle \mathbf{r}, \mathbf{s} \rangle} (1 - \mathbf{e}^{-\langle \mathbf{j}, \mathbf{s} \rangle}) \eta'_{t_0}(\mathbf{s}) d\mathbf{s} + o(\epsilon) \quad (\text{D.21})
\end{aligned}$$

Furthermore by the binomial theorem we get the next d identities

$$\begin{aligned}
(1) \quad & \sum_{i=1}^d \sum_{j=1}^q \binom{q}{j} (-1)^j (1 - \mathbf{e}^{-j\mathbf{s}}) = - \sum_{i=1}^d (1 - \mathbf{e}^{-s})^q \\
(2) \quad & \sum_{\substack{i_1, i_2 \in \{1, \dots, d\} \\ i_1 < i_2}} \sum_{j_1=1}^{q_{i_1}} \sum_{j_2=1}^{q_{i_2}} \binom{q_{i_1}}{j_1} \binom{q_{i_2}}{j_2} (-1)^{j_1+j_2} (1 - \mathbf{e}^{-j_1 s_{i_1} - j_2 s_{i_2}}) \\
&= \sum_{\substack{i_1, i_2 \in \{1, \dots, d\} \\ i_1 < i_2}} \left\{ (1 - \mathbf{e}^{-s_{i_1}})^{q_{i_1}} + (1 - \mathbf{e}^{-s_{i_2}})^{q_{i_2}} - (1 - \mathbf{e}^{-s_{i_1}})^{q_{i_1}} (1 - \mathbf{e}^{-s_{i_2}})^{q_{i_2}} \right\} \\
&\quad \vdots \\
(\text{d-1}) \quad & \sum_{\substack{i_1, \dots, i_{d-1} \in \{1, \dots, d\} \\ i_1 < \dots < i_{d-1}}} \sum_{j_1=1}^{q_{i_1}} \cdots \sum_{j_{d-1}=1}^{q_{i_{d-1}}} \binom{q_{i_1}}{j_1} \cdots \binom{q_{i_{d-1}}}{j_{d-1}} (-1)^{j_1+\cdots+j_{d-1}} (1 - \mathbf{e}^{-j_1 s_{i_1} - \cdots - j_{d-1} s_{i_{d-1}}}) \\
&= \sum_{\substack{i_1, \dots, i_{d-1} \in \{1, \dots, d\} \\ i_1 < \dots < i_{d-1}}} \left\{ (-1)^{d-1} \sum_{j=1}^{d-1} (1 - \mathbf{e}^{-s_{i_j}})^{q_{i_j}} + \right. \\
&\quad \left. (-1)^{d-2} \sum_{\substack{j_1, j_2 \in \{i_1, \dots, i_{d-1}\} \\ j_1 < j_2}} (1 - \mathbf{e}^{-s_{j_1}})^{q_{j_1}} (1 - \mathbf{e}^{-s_{j_2}})^{q_{j_2}} + \cdots - (1 - \mathbf{e}^{-s_{i_1}})^{q_{i_1}} \cdots (1 - \mathbf{e}^{-s_{i_{d-1}}})^{q_{i_{d-1}}} \right\} \\
(\text{d}) \quad & \sum_{j_1=1}^{q_1} \cdots \sum_{j_d=1}^{q_d} \binom{q_1}{j_1} \cdots \binom{q_d}{j_d} (-1)^{\langle \mathbf{1}, \mathbf{j} \rangle} (1 - \mathbf{e}^{-\langle \mathbf{j}, \mathbf{s} \rangle}) = (-1)^d \sum_{j=1}^d (1 - \mathbf{e}^{-s_j})^{q_j} + \\
&\quad (-1)^{d-1} \sum_{\substack{j_1, j_2 \in \{1, \dots, d\} \\ j_1 < j_2}} (1 - \mathbf{e}^{-s_{j_1}})^{q_{j_1}} (1 - \mathbf{e}^{-s_{j_2}})^{q_{j_2}} + \cdots - (1 - \mathbf{e}^{-s_{i_1}})^{q_{i_1}} \cdots (1 - \mathbf{e}^{-s_{i_d}})^{q_{i_d}}
\end{aligned}$$

So we have that (D.20) becomes

$$\begin{aligned}
& \mathbf{e}^{-\Delta_{t_0-\epsilon}^{t_0} \psi_t(\mathbf{r})} \left\{ 1 + \sum_{i=1}^d \sum_{j=1}^{q_i} \binom{q_i}{j} (-1)^j - \epsilon \int_{(\mathbb{R}^+)^d} \mathbf{e}^{-\langle \mathbf{r}, \mathbf{s} \rangle} \sum_{i=1}^d \sum_{j=1}^{q_i} \binom{q_i}{j} (-1)^j (1 - \mathbf{e}^{-j_1 s_1}) \eta'_{t_0}(\mathbf{s}) d\mathbf{s} \right. \\
& + \sum_{\substack{i_1, i_2 \in \{1, \dots, d\} \\ i_1 < i_2}} \sum_{j_1=1}^{q_{i_1}} \sum_{j_2=1}^{q_{i_2}} \binom{q_{i_1}}{j_1} \binom{q_{i_2}}{j_2} (-1)^{j_1+j_2} \\
& - \epsilon \int_{(\mathbb{R}^+)^d} \mathbf{e}^{-\langle \mathbf{r}, \mathbf{s} \rangle} \sum_{\substack{i_1, i_2 \in \{1, \dots, d\} \\ i_1 < i_2}} \sum_{j_1=1}^{q_{i_1}} \sum_{j_2=1}^{q_{i_2}} \binom{q_{i_1}}{j_1} \binom{q_{i_2}}{j_2} (-1)^{j_1+j_2} (1 - \mathbf{e}^{-j_1 s_{i_1} - j_2 s_{i_2}}) \eta'_{t_0}(\mathbf{s}) d\mathbf{s} \\
& + \dots + \sum_{j_1=1}^{q_1} \dots \sum_{j_d=1}^{q_d} \binom{q_1}{j_1} \dots \binom{q_d}{j_d} (-1)^{\langle \mathbf{1}, \mathbf{j} \rangle} \\
& \left. - \epsilon \int_{(\mathbb{R}^+)^d} \mathbf{e}^{-\langle \mathbf{r}, \mathbf{s} \rangle} \sum_{j_1=1}^{q_1} \dots \sum_{j_d=1}^{q_d} \binom{q_1}{j_1} \dots \binom{q_d}{j_d} (-1)^{\langle \mathbf{1}, \mathbf{j} \rangle} (1 - \mathbf{e}^{-\langle \mathbf{j}, \mathbf{s} \rangle}) \eta'_{t_0}(\mathbf{s}) d\mathbf{s} + o(\epsilon) \right\} \\
& = \mathbf{e}^{-\Delta_{t_0-\epsilon}^{t_0} \psi_t(\mathbf{r})} \left\{ \epsilon \int_{(\mathbb{R}^+)^d} \mathbf{e}^{-\langle \mathbf{r}, \mathbf{s} \rangle} (1 - \mathbf{e}^{-s_1})^{q_1} \dots (1 - \mathbf{e}^{-s_d})^{q_d} \eta'_{t_0}(\mathbf{s}) d\mathbf{s} + o(\epsilon) \right\} \\
& = \{1 + o(1)\} \left\{ \epsilon \int_{(\mathbb{R}^+)^d} \mathbf{e}^{-\langle \mathbf{r}, \mathbf{s} \rangle} (1 - \mathbf{e}^{-s_1})^{q_1} \dots (1 - \mathbf{e}^{-s_d})^{q_d} \eta'_{t_0}(\mathbf{s}) d\mathbf{s} + o(\epsilon) \right\} \\
& = \left\{ \epsilon \int_{(\mathbb{R}^+)^d} \mathbf{e}^{-\langle \mathbf{r}, \mathbf{s} \rangle} (1 - \mathbf{e}^{-s_1})^{q_1} \dots (1 - \mathbf{e}^{-s_d})^{q_d} \eta'_{t_0}(\mathbf{s}) d\mathbf{s} + o(\epsilon) \right\}
\end{aligned}$$

□

Define

$$\Gamma_{\mathbf{D}, \epsilon} = \bigcap_{i=1}^d \bigcap_{j=1}^k \left\{ ((t_1^{(i)}, \delta_1^{(i)}, \dots, t_{n_1}^{(i)}, \delta_{n_1}^{(i)}) : m_i^c(\{T_{(j)}\}) = n_{i,j}^c, m_i^e((T_{(j)} - \epsilon, T_{(j)}]) = n_{i,j}^e \right\}$$

so that

$$\mathbb{E} \left[\mathbf{e}^{-\lambda_1 \mu_1(0,t] - \dots - \lambda_d \mu_d(0,t]} | \mathbf{D} \right] = \lim_{\epsilon \rightarrow 0} \frac{\mathbb{E} \left[\mathbf{e}^{-\lambda_1 \mu_1(0,t] - \dots - \lambda_d \mu_d(0,t]} \mathbb{1}_{\Gamma_{\mathbf{D}, \epsilon}}(\mathbf{D}) \right]}{\mathbb{P}[\mathbf{D} \in \Gamma_{\mathbf{D}, \epsilon}]}$$

We observe that defining $T_{(0)} = 0$, $\bar{n}_{i,k+1}^e = 0$ for $i \in \{1, \dots, d\}$ and selecting ϵ sufficiently small such that $t \notin (T_{(j)} - \epsilon, T_{(j)})$ for all $j \in \{1, \dots, k\}$

$$\begin{aligned}
& \mathbb{E} \left[\mathbf{e}^{-\lambda_1 \mu_1(0,t] - \dots - \lambda_d \mu_d(0,t]} \mathbb{1}_{\Gamma_{\mathbf{D},\epsilon}}(\mathbf{D}) \mid (\mu_1, \dots, \mu_d) \right] \\
&= \prod_{i=1}^d e^{-\lambda_i \mu_i(0,t]} \prod_{j=1}^k \mathbf{e}^{-n_{i,j}^c \mu_i(0,T_{(j)}) - n_{i,j}^e \mu_i(0,T_{(j)} - \epsilon]} \left(1 - \mathbf{e}^{-\mu_i(T_{(j)} - \epsilon, T_{(j)})} \right)^{n_{i,j}^e} \\
&= \prod_{i=1}^d e^{-\lambda_i \mathbb{1}_{(0,t]}(T_{(k)}) \mu_i(T_{(k)}, t]} \prod_{j=1}^k \left\{ \mathbf{e}^{-\lambda_i \mathbb{1}_{(0,t)}(T_{(j-1)}) \mu_i(T_{(j-1)}, \min\{t, T_{(j)} - \epsilon\}] - \lambda_i \mathbb{1}_{(0,t]}(T_{(j)}) \mu_i(T_{(j)} - \epsilon, T_{(j)})} \right. \\
&\quad \times \mathbf{e}^{-n_{i,j}^c \sum_{r=1}^j (\mu_i(T_r - \epsilon, T_r)] + \mu_i(T_{(r-1)}, T_r) - \epsilon]} - n_{i,j}^e \sum_{r=1}^j \mu_i(T_{(r-1)}, T_r) - \epsilon - n_{i,j}^e \sum_{r=1}^{j-1} \mu_i(T_r) - \epsilon, T_r] \\
&\quad \times \left(1 - \mathbf{e}^{-\mu_i(T_{(j)} - \epsilon, T_{(j)})} \right)^{n_{i,j}^e} \Big\} \\
&= \prod_{i=1}^d \left\{ \mathbf{e}^{-\lambda_i \mathbb{1}_{(0,t)}(T_{(k)}) \mu_i(T_{(k)}, t] - \sum_{j=1}^k n_{i,j}^c \sum_{r=1}^j (\mu_i(T_r - \epsilon, T_r)] + \mu_i(T_{(r-1)}, T_r) - \epsilon]} \right. \\
&\quad \times \mathbf{e}^{-\sum_{j=1}^k n_{i,j}^e \sum_{r=1}^j \mu_i(T_{(r-1)}, T_r) - \epsilon - \sum_{j=1}^k n_{i,j}^e \sum_{r=1}^{j-1} \mu_i(T_r) - \epsilon, T_r]} \\
&\quad \times \prod_{j=1}^k \left\{ \mathbf{e}^{-\lambda_i \mathbb{1}_{(0,t)}(T_{(j-1)}) \mu_i(T_{(j-1)}, \min\{t, T_{(j)} - \epsilon\}] - \lambda_i \mathbb{1}_{(0,t]}(T_{(j)}) \mu_i(T_{(j)} - \epsilon, T_{(j)})} \left(1 - \mathbf{e}^{-\mu_i(T_{(j)} - \epsilon, T_{(j)})} \right)^{n_{i,j}^e} \right\} \Big\} \\
&= \prod_{i=1}^d \left\{ \prod_{j=1}^k \left\{ \mathbf{e}^{-[\lambda_i \mathbb{1}_{(0,t]}(T_{(j)}) + \bar{n}_{i,j}^c + \bar{n}_{i,j+1}^e] \mu_i(T_{(j)} - \epsilon, T_{(j)})} \left(1 - \mathbf{e}^{-\mu_i(T_{(j)} - \epsilon, T_{(j)})} \right)^{n_{i,j}^e} \right\} e^{-\lambda_i \mathbb{1}_{(0,t]}(T_{(k)}) \mu_i(T_{(k)}, t]} \right. \\
&\quad \times \prod_{j=1}^k \left\{ \mathbf{e}^{-\lambda_i \mathbb{1}_{(0,t)}(T_{(j-1)}) \mu_i(T_{(j-1)}, \min\{t, T_{(j)} - \epsilon\}] - \bar{n}_{i,j}^c \mu_i(T_{(j-1)}, T_{(j)} - \epsilon] - \bar{n}_{i,j}^e \mu_i(T_{(j-1)}, T_{(j)} - \epsilon]} \right\} \Big\}
\end{aligned}$$

So defining

$$\begin{aligned}
I_{1,\epsilon} &= \prod_{j=1}^k \prod_{i=1}^d \left\{ \mathbf{e}^{-[\lambda_i \mathbb{1}_{(0,t]}(T_{(j)}) + \bar{n}_{i,j}^c + \bar{n}_{i,j+1}^e] \mu_i(T_{(j)} - \epsilon, T_{(j)})} \left(1 - \mathbf{e}^{-\mu_i(T_{(j)} - \epsilon, T_{(j)})} \right)^{n_{i,j}^e} \right\} \\
I_{2,\epsilon} &= \prod_{i=1}^d e^{-\lambda_i \mathbb{1}_{(0,t]}(T_{(k)}) \mu_i(T_{(k)}, t]} \prod_{j=1}^k \left\{ \mathbf{e}^{-\lambda_i \mathbb{1}_{(0,t)}(T_{(j-1)}) \mu_i(T_{(j-1)}, \min\{t, T_{(j)} - \epsilon\}] - (\bar{n}_{i,j}^c + \bar{n}_{i,j}^e) \mu_i(T_{(j-1)}, T_{(j)} - \epsilon]} \right\}
\end{aligned}$$

We get from the independence property of CRM's that

$$\mathbb{E} \left[\mathbf{e}^{-\lambda_1 \mu_1(0,t] - \dots - \lambda_d \mu_d(0,t]} \mathbb{1}_{\Gamma_{\mathbf{D},\epsilon}}(\mathbf{D}) \right] = \mathbb{E}[I_{1,\epsilon}] \mathbb{E}[I_{2,\epsilon}] \quad (\text{D.22})$$

We observe that for $r_i = \lambda_i \mathbb{1}_{(0,t]}(T_{(j)}) + \bar{n}_{i,j}^c + \bar{n}_{i,j+1}^e$, $i \in \{1, \dots, d\}$ we have that $\min\{r_1, \dots, r_d\} \geq 1$ and for $j \in \{1, \dots, k\}$ such that $T_{(j)}$ is an exact observation we have that $\max\{n_{1,j}, \dots, n_{d,j}\} \geq 1$ so lemma 2 can be applied yielding

$$\begin{aligned}
& \mathbb{E} \left[\prod_{i=1}^d \left\{ \mathbf{e}^{-[\lambda_i \mathbb{1}_{(0,t]}(T_{(j)}) + \bar{n}_{i,j}^c + \bar{n}_{i,j+1}^e] \mu_i(T_{(j)} - \epsilon, T_{(j)})} \left(1 - \mathbf{e}^{-\mu_i(T_{(j)} - \epsilon, T_{(j)})} \right)^{n_{i,j}^e} \right\} \right] \\
&= \epsilon \int_{(\mathbb{R}^+)^d} \prod_{i=1}^d \left\{ \mathbf{e}^{-[\lambda_i \mathbb{1}_{(0,t]}(T_{(j)}) + \bar{n}_{i,j}^c + \bar{n}_{i,j+1}^e] s_i} (1 - \mathbf{e}^{-s_i})^{n_{i,j}^e} \right\} \eta'_t T_{(j)}(\mathbf{s}) d\mathbf{s} + o(\epsilon) \quad (\text{D.23})
\end{aligned}$$

On the other hand, for $j \notin \mathcal{J} = \{j : T_{(j)} \text{ is an exact observation}\}$ we have $n_{i,j}^e = 0$ so by the continuity

of $\eta_t(\mathbf{s})$ in t we have

$$\begin{aligned} & \lim_{\epsilon \rightarrow 0} \mathbb{E} \left[\prod_{i=1}^d \left\{ \mathbf{e}^{-[\lambda_i \mathbb{1}_{(0,t]}(T_{(j)} + \bar{n}_{i,j}^\epsilon + \bar{n}_{i,j+1}^\epsilon)] \mu_i(T_{(j)} - \epsilon, T_{(j)})} \left(1 - \mathbf{e}^{-\mu_i(T_{(j)} - \epsilon, T_{(j)})} \right)^{n_{i,j}^\epsilon} \right\} \right] \\ &= \lim_{\epsilon \rightarrow 0} \mathbb{E} \left[\prod_{i=1}^d \left\{ \mathbf{e}^{-[\lambda_i \mathbb{1}_{(0,t]}(T_{(j)} + \bar{n}_{i,j}^\epsilon + \bar{n}_{i,j+1}^\epsilon)] \mu_i(T_{(j)} - \epsilon, T_{(j)})} \right\} \right] = 1 \end{aligned} \quad (\text{D.24})$$

From (D.23), (D.24) and the independence property of CRM's we obtain

$$\lim_{\epsilon \rightarrow 0} \mathbb{E}[I_{1,\epsilon}] = \lim_{\epsilon \rightarrow 0} \prod_{j \in \mathcal{J}} \left\{ \epsilon \int_{(\mathbb{R}^+)^d} \prod_{i=1}^d \left\{ \mathbf{e}^{-[\lambda_i \mathbb{1}_{(0,t]}(T_{(j)}) + \bar{n}_{i,j}^\epsilon + \bar{n}_{i,j+1}^\epsilon] s_i} (1 - \mathbf{e}^{-s_i})^{n_{i,j}^\epsilon} \right\} \eta'_{T_{(j)}}(\mathbf{s}) d\mathbf{s} + o(\epsilon) \right\}$$

Also by continuity and independence, defining $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_d)$, we get

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \mathbb{E}[I_{2,\epsilon}] &= \mathbf{e}^{-[\psi_t(\mathbb{1}_{(0,t]}(T_{(k)})\boldsymbol{\lambda}) - \psi_{T_{(k)}}(\mathbb{1}_{(0,t]}(T_{(k)})\boldsymbol{\lambda})]} \times \\ &\times \prod_{j=1}^k \left\{ \mathbf{e}^{-[\psi_{t \wedge T_{(j)}}(\mathbb{1}_{(0,t]}(T_{(j-1)})\boldsymbol{\lambda} + \bar{\mathbf{n}}_j^\epsilon + \bar{\mathbf{n}}_j^\epsilon) - \psi_{T_{(j-1)}}(\mathbb{1}_{(0,t]}(T_{(j-1)})\boldsymbol{\lambda} + \bar{\mathbf{n}}_j^\epsilon + \bar{\mathbf{n}}_j^\epsilon)] - [\psi_{T_{(j)}}(\bar{\mathbf{n}}_j^\epsilon + \bar{\mathbf{n}}_j^\epsilon) - \psi_{t \wedge T_{(j)}}(\bar{\mathbf{n}}_j^\epsilon + \bar{\mathbf{n}}_j^\epsilon)]} \right\} \end{aligned}$$

So by (D.22), (D.24) and (D.23) we get that

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \mathbb{E} \left[\mathbf{e}^{-\lambda_1 \mu_1(0,t] - \dots - \lambda_d \mu_d(0,t]} \mathbb{1}_{\Gamma_{D,\epsilon}}(\mathbf{D}) \right] &= \mathbf{e}^{-\Delta_{T_{(k)}}^t \psi_t(\mathbb{1}_{(0,t]}(T_{(k)})\boldsymbol{\lambda}) - \sum_{j=1}^k \Delta_{T_{(j-1)}}^{t \wedge T_{(j)}} \psi_t(\mathbb{1}_{(0,t]}(T_{(j-1)})\boldsymbol{\lambda} + \bar{\mathbf{n}}_j^\epsilon + \bar{\mathbf{n}}_j^\epsilon)} \\ &\times \prod_{j \in \mathcal{J}} \lim_{\epsilon \rightarrow 0} \left\{ \epsilon \int_{(\mathbb{R}^+)^d} \prod_{i=1}^d \left\{ \mathbf{e}^{-[\lambda_i \mathbb{1}_{(0,t]}(T_{(j)}) + \bar{n}_{i,j}^\epsilon + \bar{n}_{i,j+1}^\epsilon] s_i} (1 - \mathbf{e}^{-s_i})^{n_{i,j}^\epsilon} \right\} \eta'_{T_{(j)}}(\mathbf{s}) d\mathbf{s} + o(\epsilon) \right\} \\ &\times \mathbf{e}^{-\sum_{j=1}^k \Delta_{t \wedge T_{(j)}}^{T_{(j)}} \psi_t(\bar{\mathbf{n}}_j^\epsilon + \bar{\mathbf{n}}_j^\epsilon)} \end{aligned}$$

And similarly

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \mathbb{P}[\mathbf{D} \in \Gamma_{D,\epsilon}] &= \mathbf{e}^{-\sum_{j=1}^k \Delta_{T_{(j-1)}}^{T_{(j)}} \psi_t(\bar{\mathbf{n}}_j^\epsilon + \bar{\mathbf{n}}_j^\epsilon)} \\ &\times \prod_{j \in \mathcal{J}} \lim_{\epsilon \rightarrow 0} \left\{ \epsilon \int_{(\mathbb{R}^+)^d} \prod_{i=1}^d \left\{ \mathbf{e}^{-[\bar{n}_{i,j}^\epsilon + \bar{n}_{i,j+1}^\epsilon] s_i} (1 - \mathbf{e}^{-s_i})^{n_{i,j}^\epsilon} \right\} \eta'_{T_{(j)}}(\mathbf{s}) d\mathbf{s} + o(\epsilon) \right\} \end{aligned}$$

We set $T_{(k+1)} = \infty$ so we conclude

$$\begin{aligned} \mathbb{E} \left[\mathbf{e}^{-\lambda_1 \mu_1(0,t] - \dots - \lambda_d \mu_d(0,t]} | \mathbf{D} \right] &= \lim_{\epsilon \rightarrow 0} \frac{\mathbb{E} \left[\mathbf{e}^{-\lambda_1 \mu_1(0,t] - \dots - \lambda_d \mu_d(0,t]} \mathbb{1}_{\Gamma_{D,\epsilon}}(\mathbf{D}) \right]}{\mathbb{P}[\mathbf{D} \in \Gamma_{D,\epsilon}]} \\ &= \mathbf{e}^{-\sum_{j=1}^{k+1} \Delta_{T_{(j-1)}}^{t \wedge T_{(j)}} [\psi_t(\mathbb{1}_{(0,t]}(T_{(j-1)})\boldsymbol{\lambda} + \bar{\mathbf{n}}_j^\epsilon + \bar{\mathbf{n}}_j^\epsilon) - \psi_t(\bar{\mathbf{n}}_j^\epsilon + \bar{\mathbf{n}}_j^\epsilon)]} \\ &\times \prod_{j \in \mathcal{J}} \lim_{\epsilon \rightarrow 0} \left\{ \frac{\epsilon \int_{(\mathbb{R}^+)^d} \prod_{i=1}^d \left\{ \mathbf{e}^{-[\lambda_i \mathbb{1}_{(0,t]}(T_{(j)}) + \bar{n}_{i,j}^\epsilon + \bar{n}_{i,j+1}^\epsilon] s_i} (1 - \mathbf{e}^{-s_i})^{n_{i,j}^\epsilon} \right\} \eta'_{T_{(j)}}(\mathbf{s}) d\mathbf{s} + o(\epsilon)}{\epsilon \int_{(\mathbb{R}^+)^d} \prod_{i=1}^d \left\{ \mathbf{e}^{-[\bar{n}_{i,j}^\epsilon + \bar{n}_{i,j+1}^\epsilon] s_i} (1 - \mathbf{e}^{-s_i})^{n_{i,j}^\epsilon} \right\} \eta'_{T_{(j)}}(\mathbf{s}) d\mathbf{s} + o(\epsilon)} \right\} \\ &= \mathbf{e}^{-\sum_{j=1}^{k+1} \int_{(\mathbb{R}^+)^d \times (T_{(j-1)}, t \wedge T_{(j)}]} \mathbb{1}_{(0,t]}(T_{(j-1)}) (1 - \mathbf{e}^{-\langle \boldsymbol{\lambda}, \mathbf{s} \rangle}) \mathbf{e}^{-\langle \bar{\mathbf{n}}_j^\epsilon + \bar{\mathbf{n}}_j^\epsilon, \mathbf{s} \rangle} \nu(d\mathbf{s}, du)} \\ &\times \prod_{j \in \mathcal{J}} \left\{ \frac{\int_{(\mathbb{R}^+)^d} \prod_{i=1}^d \left\{ \mathbf{e}^{-[\lambda_i \mathbb{1}_{(0,t]}(T_{(j)}) + \bar{n}_{i,j}^\epsilon + \bar{n}_{i,j+1}^\epsilon] s_i} (1 - \mathbf{e}^{-s_i})^{n_{i,j}^\epsilon} \right\} \eta'_{T_{(j)}}(\mathbf{s}) d\mathbf{s}}{\int_{(\mathbb{R}^+)^d} \prod_{i=1}^d \left\{ \mathbf{e}^{-[\bar{n}_{i,j}^\epsilon + \bar{n}_{i,j+1}^\epsilon] s_i} (1 - \mathbf{e}^{-s_i})^{n_{i,j}^\epsilon} \right\} \eta'_{T_{(j)}}(\mathbf{s}) d\mathbf{s}} \right\} \end{aligned}$$

References

- [1] Aalen, O., Borgan, O., and Gjessing, H. (2008) *Survival and event history analysis: a process point of view*. Springer Science & Business Media.
- [2] Aldous, D. J. (1985) Exchangeability and related topics (pp. 1-198). Springer Berlin Heidelberg.
- [3] Bassetti F., Casarin R. and Leisen F. (2014). Beta-Product dependent Pitman-Yor Processes for Bayesian inference. *Journal of Econometrics*. **180**, 49–72.
- [4] Cont, R. and Tankov, P. (2004) *Financial modelling with jump processes*. Chapman & Hall.
- [5] Daley, D. J., and Vere-Jones, D. (2003) *An Introduction to the Theory of Point Processes Volume I: Elementary Theory and Methods* Springer Science & Business Media.
- [6] Doksum, K. (1974) Tailfree and neutral random probabilities and their posterior distributions. *Ann. Probab.* **2**, 183-201.
- [7] Dykstra, R. L. and Laud, P. (1981) A Bayesian nonparametric approach to reliability. *The Annals of Statistics*, **9**, 356-367.
- [8] Esmaeili, H. and Klüppelberg, C. (2010) Parameter estimation of a bivariate compound Poisson process. *Insurance: mathematics and economics*, **47**, 224–233.
- [9] Esmaeili, H. and Klüppelberg, C. (2011) Parametric estimation of a bivariate stable Lévy process. *Journal of Multivariate Analysis*, **102.5**, 918–930.
- [10] Epifani, I. and Lijoi, A. (2010) Nonparametric priors for vectors of survival functions. *Statistica Sinica*, **20**, 1455–1484.
- [11] Ferguson, T. S. (1974) Prior distributions on spaces of probability measures. *The Annals of Statistics*, **4**, 615-629.
- [12] Griffin, J. and Leisen, F. (2017) Compound random measures and their use in Bayesian nonparametrics. *Journal of the Royal Statistical Society - Series B*, **79**, 525-545.
- [13] Griffin J. E., Kolossatis M. and Steel M. F. J. (2013). Comparing Distributions By Using Dependent Normalized Random-Measure Mixtures. *Journal of the Royal Statistical Society, Series B*, **75**, 499–529.
- [14] Hjort, N. L., Holmes, C., Müller, P., and Walker, S. G. (Eds.) (2010) *Bayesian nonparametrics*. Cambridge University Press, **28**.
- [15] Ishwaran, H. and Zarepour, M. (2009). Series representations for multivariate generalized gamma processes via a scale invariance principle. *Statistica Sinica*, pp.1665-1682.
- [16] Kallsen, J. and Tankov, P. (2006) Characterization of dependence of multidimensional Lévy processes using Lévy copulas. *Journal of Multivariate Analysis*, **97**, 1551–1572.
- [17] Kim, Y. and Lee, J. (2003) Bayesian analysis of proportional hazard models. *Annals of statistics*, **31**, 493-511.
- [18] Kingman, J. (1967) Completely random measures. *Pacific Journal of Mathematics*, **21**, 59–78.
- [19] Leisen, F., and Lijoi, A. (2011) Vectors of two-parameter PoissonDirichlet processes. *Journal of Multivariate Analysis*, **102**, 482-495.
- [20] Leisen F., Lijoi A. and Spano D.(2013). A Vector of Dirichlet processes. *Electronic Journal of Statistics* **7**, 62–90.
- [21] Lijoi A., and Nipoti B.(2014), ‘A class of hazard rate mixtures for combining survival data from different experiments’, *Journal of the American Statistical Association*, **109**, 802–814.

- [22] Lijoi A., Nipoti B. and Prünster I. (2014a), ‘Bayesian inference with dependent normalized completely random measures’, *Bernoulli*, **20**, 1260–1291.
- [23] Lijoi A., Nipoti B. and Prünster I. (2014b), ‘Dependent mixture models: clustering and borrowing information’, *Computational Statistics and Data Analysis*, **71**, 417–433.
- [24] MacEachern S. N. (1999). Dependent nonparametric processes. In *ASA Proceedings of the Section on Bayesian Statistical Science*, Alexandria, VA: American Statistical Association.
- [25] Müller P., Quintana F. and Rosner G. L. (2004). A method for combining inference across related nonparametric Bayesian models. *Journal of the Royal Statistical Society, Series B*, **66**, 735–749.
- [26] Nelsen, Roger B. (2013) *An introduction to copulas*. Springer Science & Business Media, **139**.
- [27] Todeschini A. and Caron F. (2016). Exchangeable random measures for sparse and modular graphs with overlapping communities. arXiv:1602.02114.
- [28] Woolson, R. F., and Lachenbruch, P. A. (1980) Rank tests for censored matched pairs. *Biometrika*, **67**, 597-606.
- [29] Zhu, W., and Leisen, F. (2015). “A multivariate extension of a vector of two-parameter PoissonDirichlet processes.” *Journal of Nonparametric Statistics*, **27**, 89-105.